

NYE Fundraising data

What have we learned this past year?

Luisa M. Mimmi

Goals of proposed analysis

1. How have the main Donation/Revenue streams evolved between NYE 2011 – NYE 2019?
2. Zooming in on *Individual(s) & Households*:
 - Any characteristics correlated to being a donor and/or \$\$ donated?
 - (Based on them) Could we predict future behavior?
3. Thoughts and suggestions

1) OVERVIEW OF DONATIONS NYE2011-NYE2019

A sense of magnitude of the available data

Upon merging with **Sendgrid** We currently have in **Salesforce**:

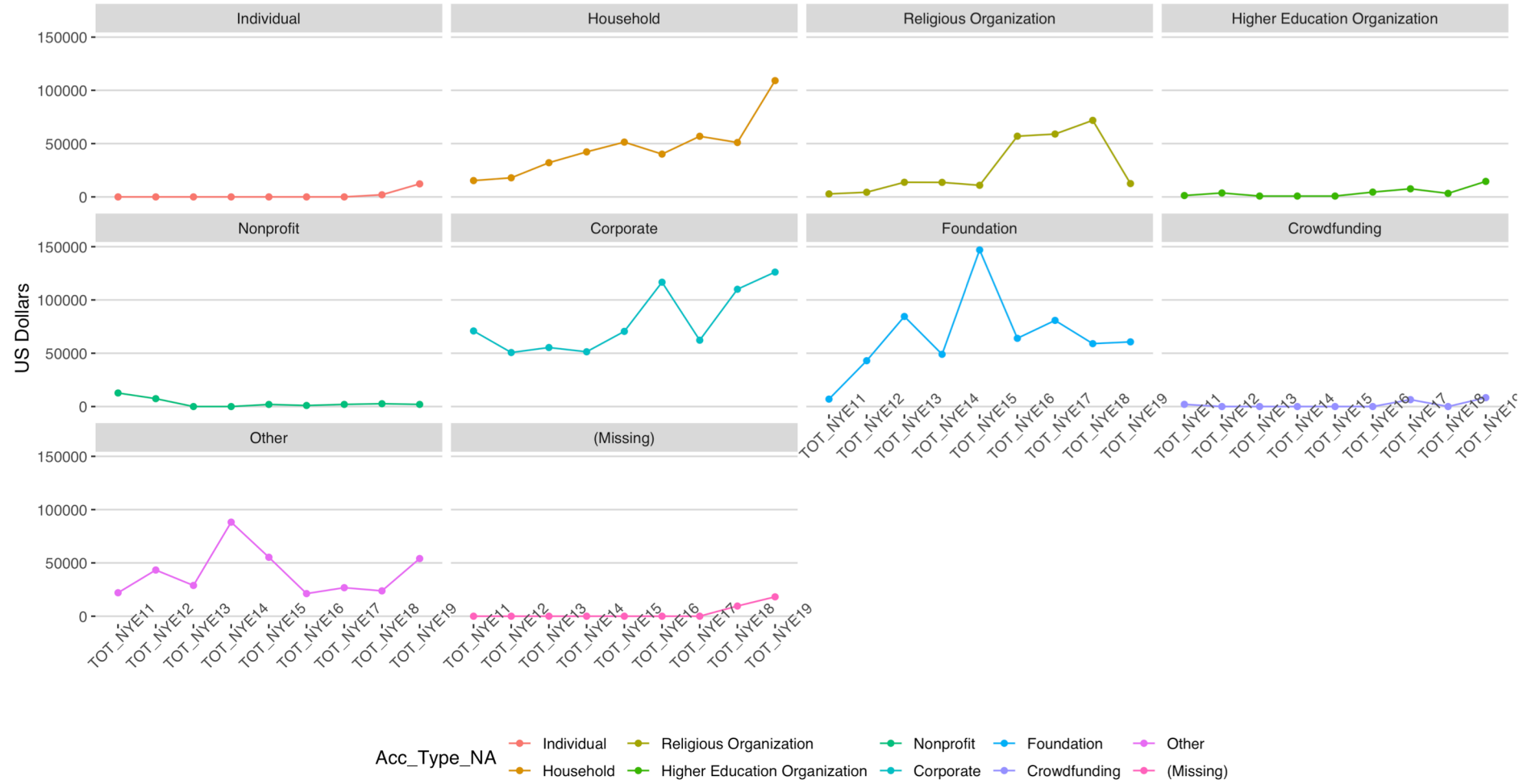
- **25,650 Contacts:**
 - 17,084 with an active email
 - of which 6,341 opened a NYE email after 2018
- (linked to) **24,447 Accounts:**
 - 398 (**1.6%**) **donated at least once** between 2011-2019

ACCOUNT TYPE	N	% of Tot Account
(Missing)	17203	70.28%
Household	4949	20.22%
Individual	1703	6.96%
Religious Organization	343	1.40%
Higher Educ Organization	97	0.40%
Corporate	70	0.29%
Press	56	0.23%
Foundation	28	0.11%
Other	17	0.07%
Nonprofit	5	0.02%
Government	4	0.02%
Crowdfunding	2	0.01%
TOTAL	24477	100.00%

Our main focus today

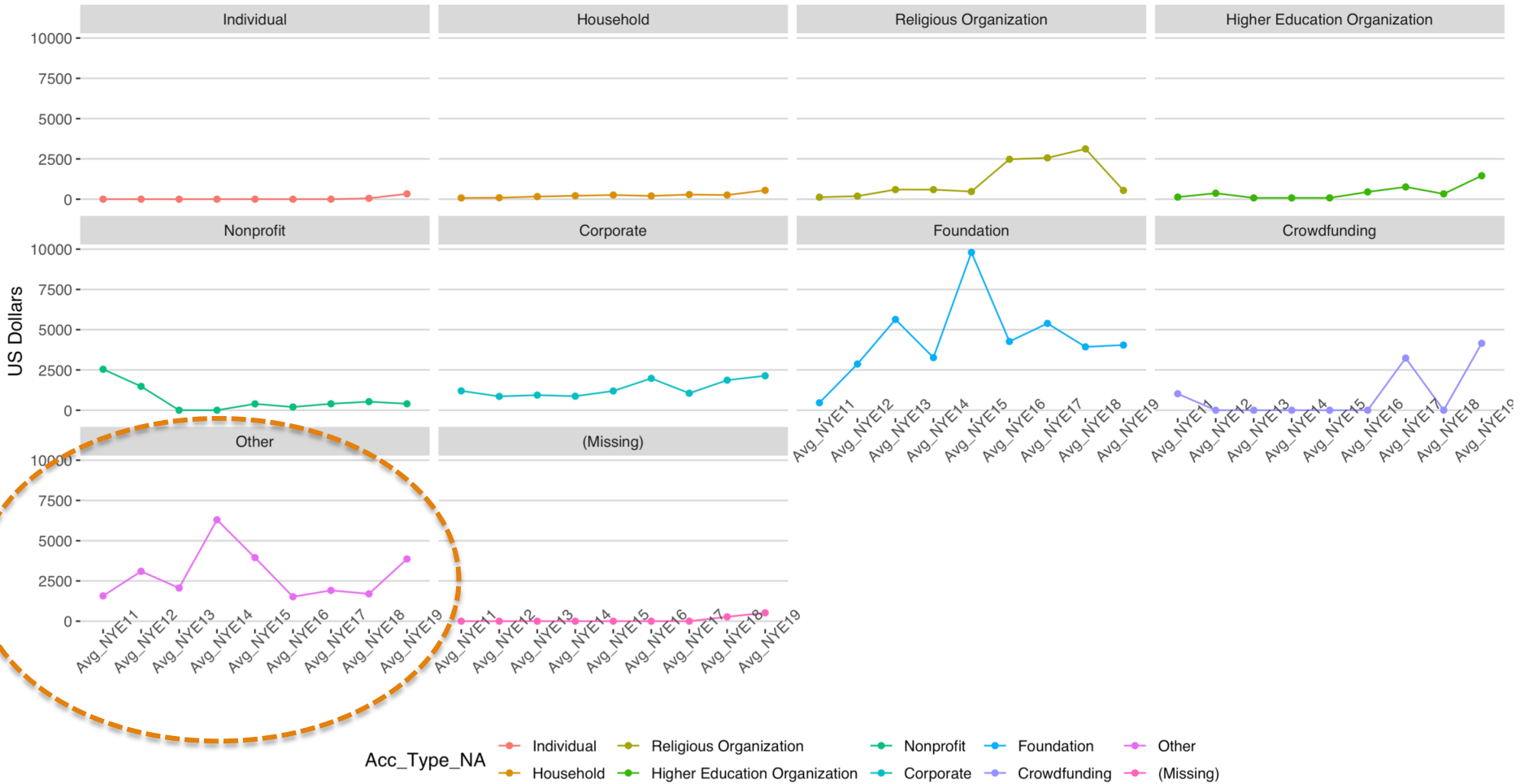
TOT Donations over years by: Type of ACCOUNT

Total donations over the years by Type of Accounts



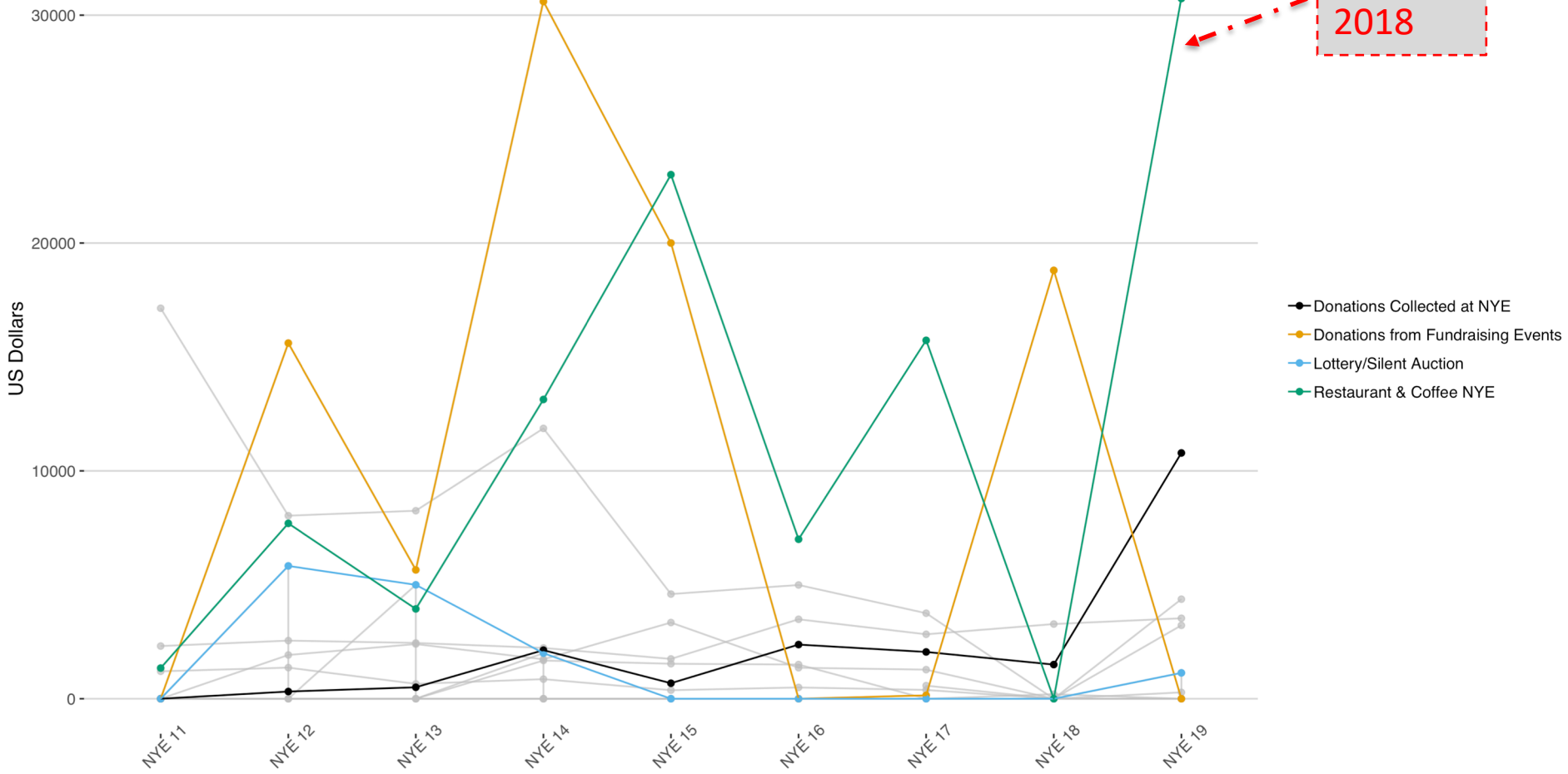
Avg. Donations over years by: Type of ACCOUNT

Average donations over the years by Type of Accounts



Breaking Down strongest Revenue streams from NYE event...

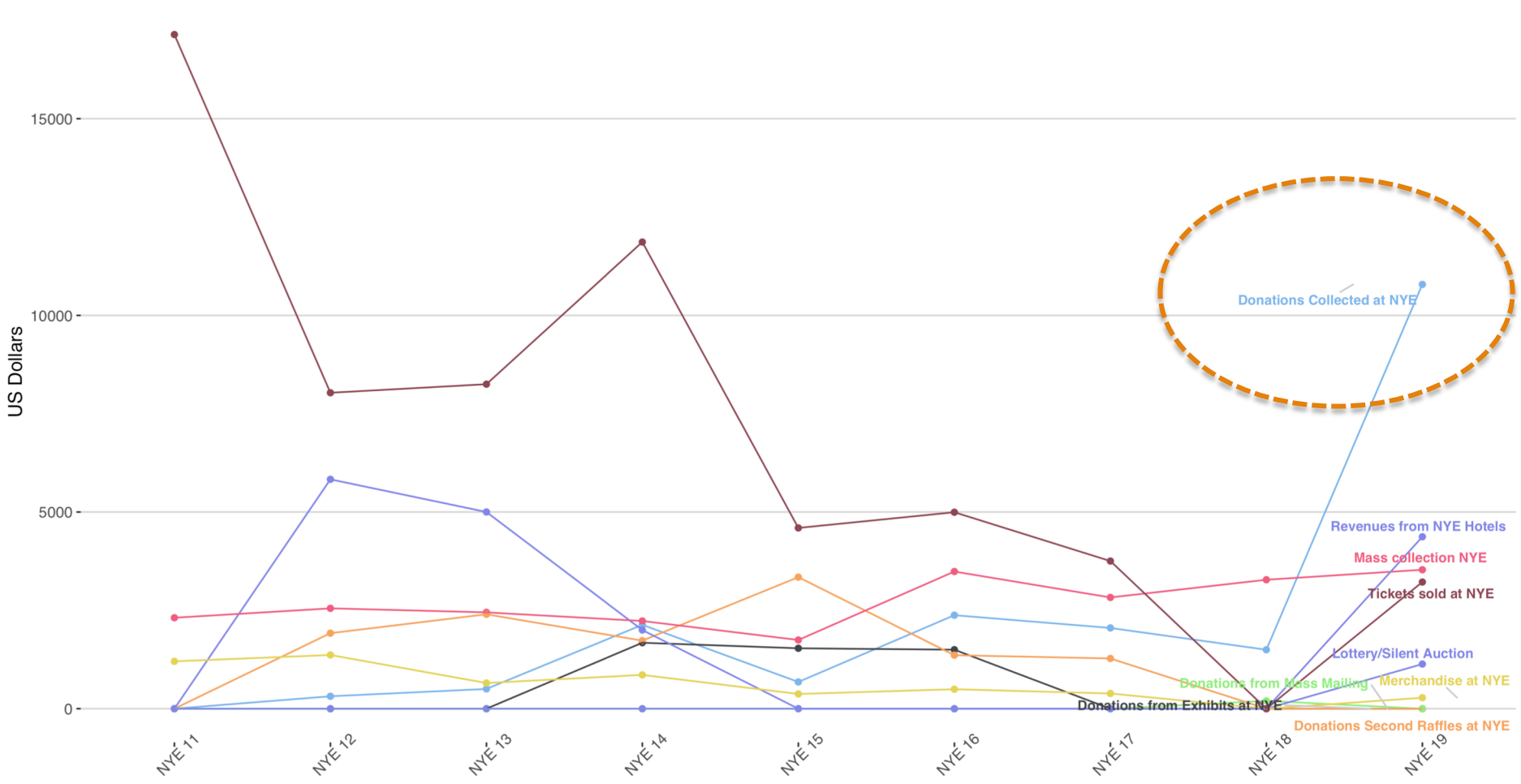
Some interesting donation streams over the years



Created on 2019-03-24

At NYE... (without Restaurant & Coffee NYE)

"Other Account" donations / revenue streams over the years



Created on 2019-03-24

Some observations on patterns of donations NYE2011-NYE2019

- 2019 Donations collected @ NYE - including Fundraising table 😊 became 2nd largest contribution!!!
- Are we making a profit from Hotel booking !?!?
- Lottery/Silent Auction... not much
- Outlier revenue from Tickets sold at NYE11!!! (what was it?)
- (Revenue from _Restaurant & Coffee_ missing NYE18 data point)

BOTTOM LINE:

these fundraising streams seem highly dependent on effort... something to think about to prioritize

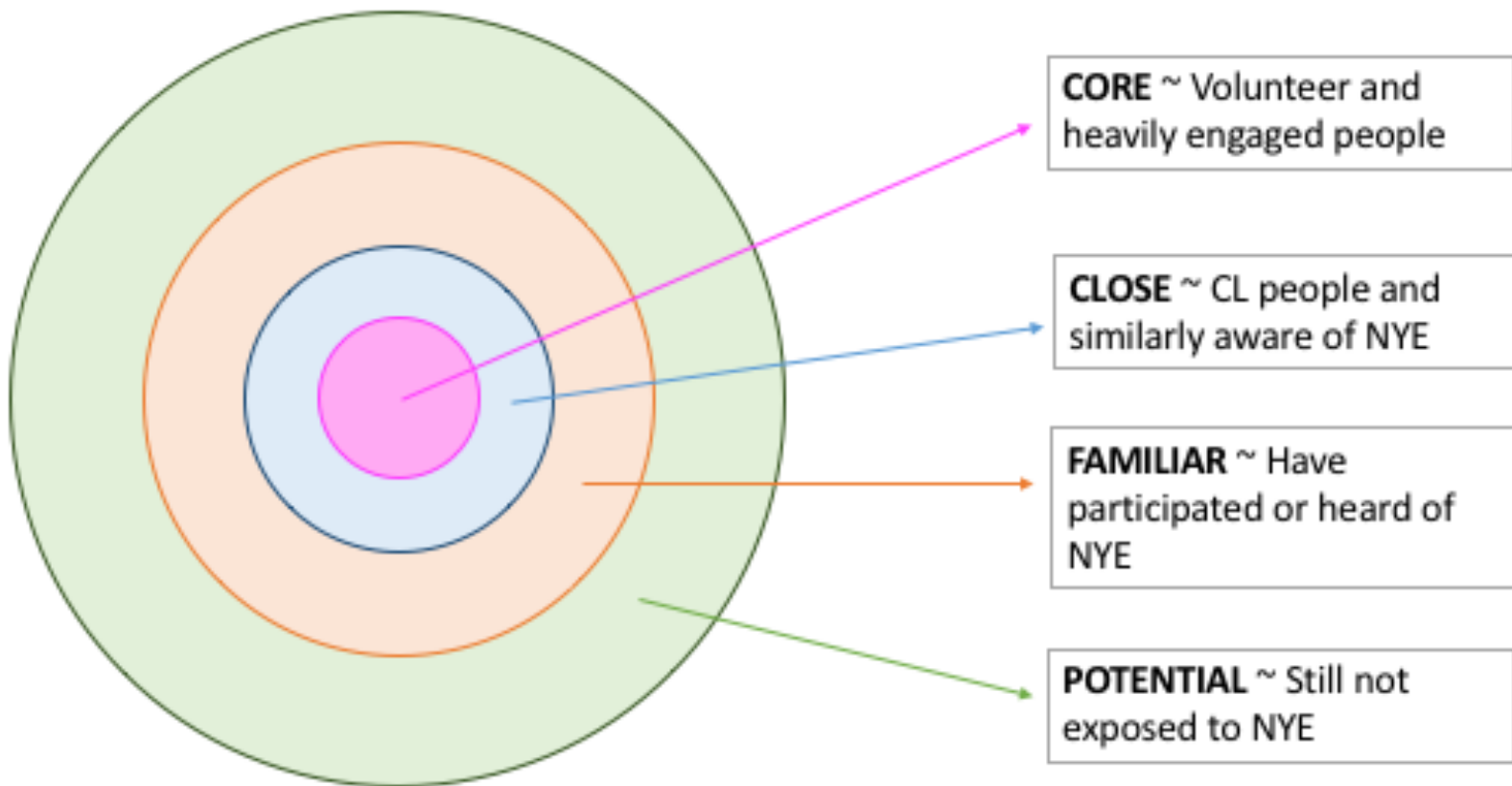
2) ZOOMING IN ON INDIVIDUAL/ HOUSEHOLD CONTRIBUTORS

Which characteristics seem associated to being a donor?

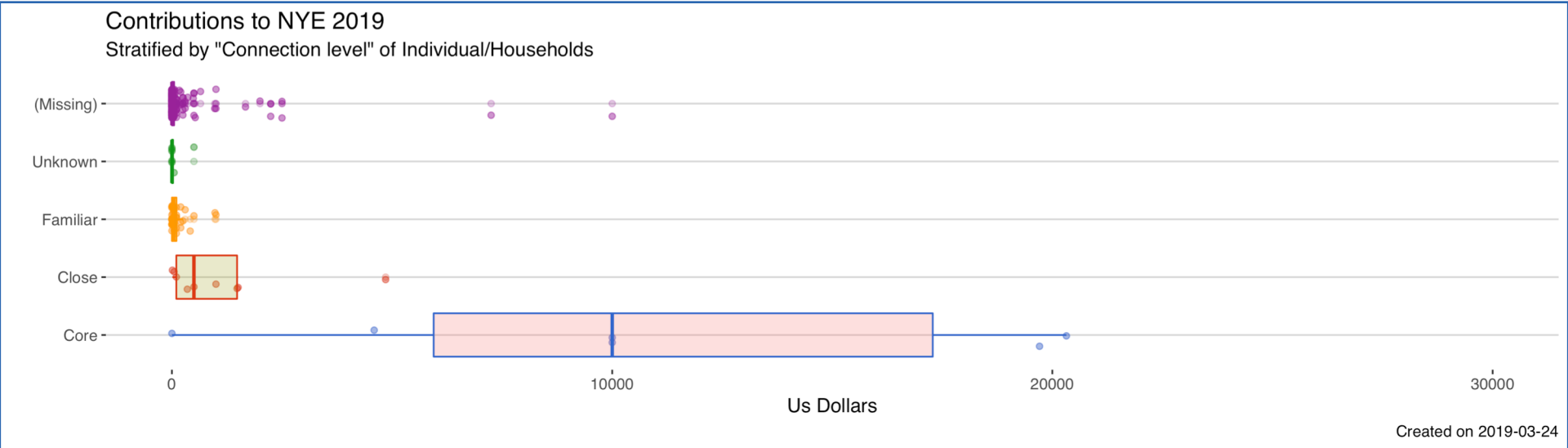
Measure of association between "Is Donor" and some Categorical Variables

Cramer's V phi statistic	Categorical Var
0.2366	Connection
0.0125	VIP status
0.0292	Speaker status
0.0767	Volunteer status
0.1429	Has Email
0.1431	Has Valid Email
0.0323	Has Valid Email but Unsubscribed
0.1193	How recently Opened Last Email

Classification of Contacts based on “CLOSENESS” to NYE

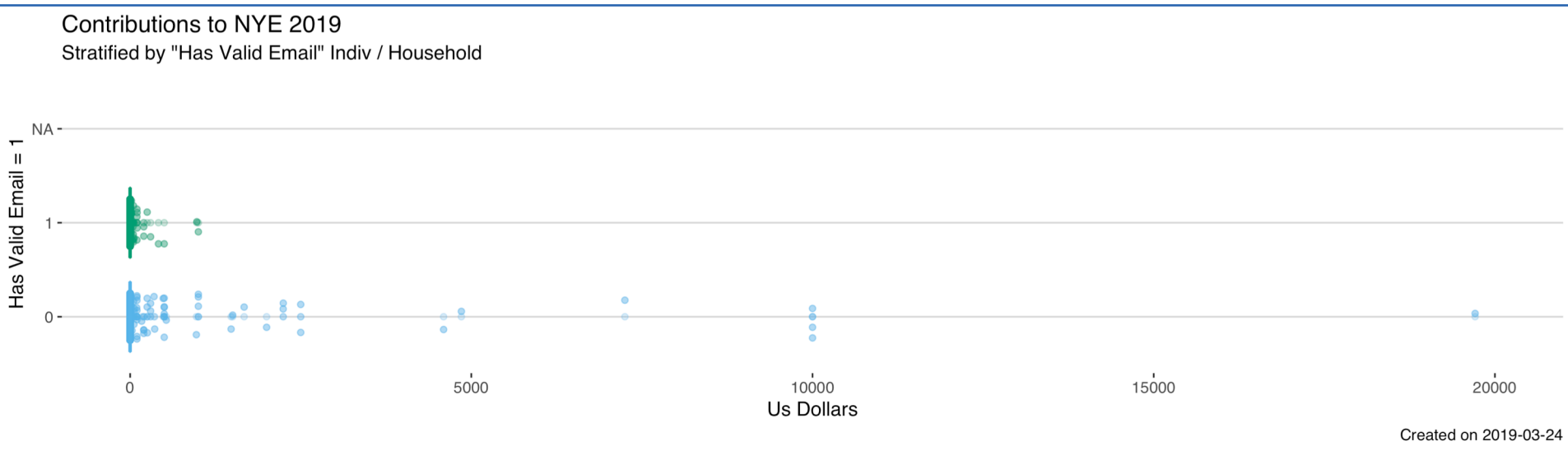


“Is being a "CLOSE" friend a driver of donation?" (YEAH!)



Few, incredibly generous and engaged CORE organizers... & CLOSE friends

How effective are our NYE email blasts? (1/2)



() Concerning impression, but caveats apply...*

How effective are our NYE email blasts? (2/2)

Contributions to NYE 2019
Stratified by "Connection level" of Individual/Households



Created on 2019-03-22

() Concerning impression, but caveats apply...*

2.B) “ATTEMPTED” PREDICTION EXPERIMENT...

I use a sub-sample of Individual/Household donors for the experiment (rebalanced*)

From the 6,552 Individual or Household Accounts, I took a:

Sample = 887

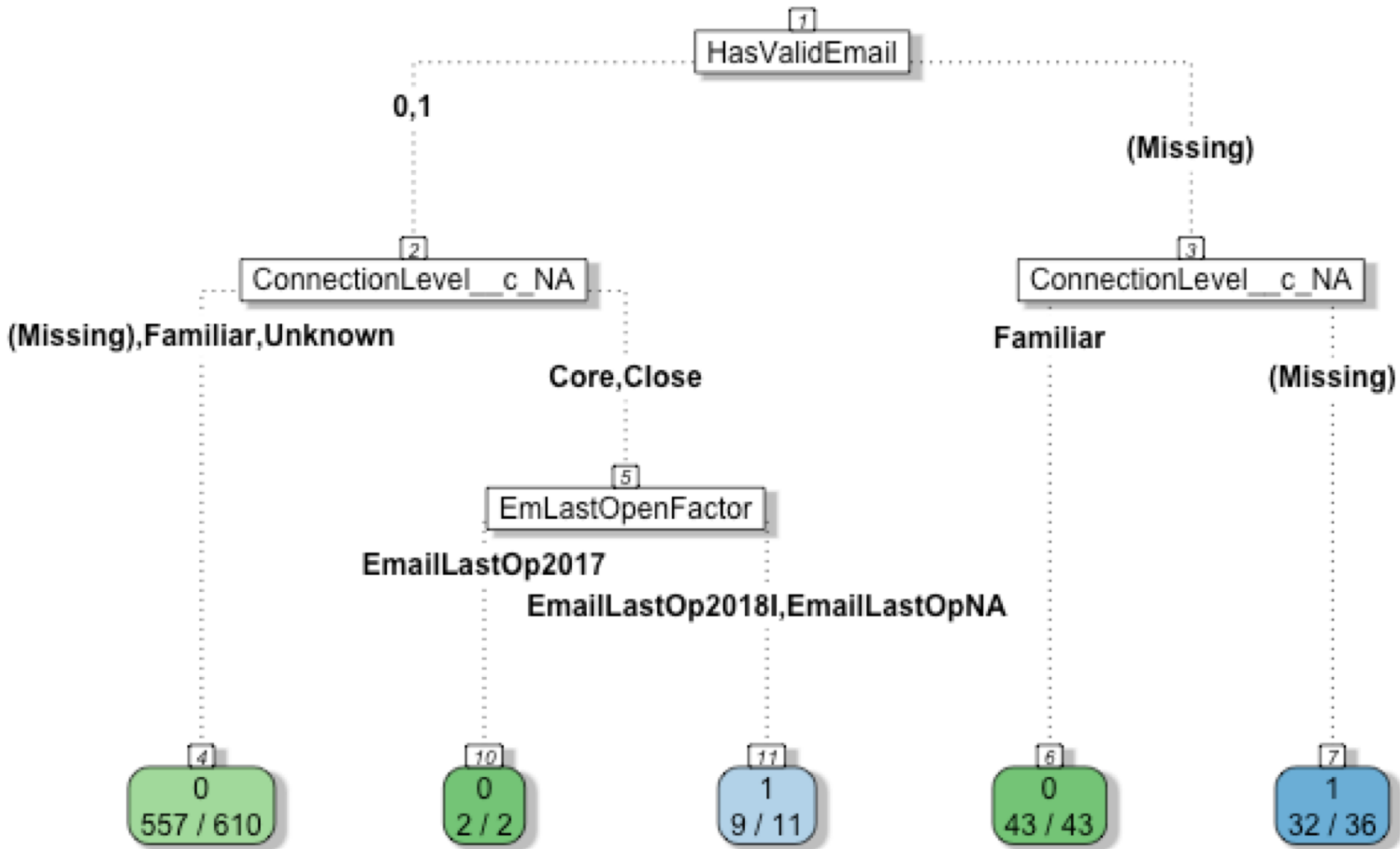
In which:

IsDonor	n	percent	HasValidEmail	n	percent
0	642	73.20%	0	378	43.10%
1	235	26.80%	1	408	46.52%
			(Missing)	91	10.38%

ConnectionLevel_c_NA	n	percent
(Missing)	376	42.9%
Core	6	0.7%
Close	12	1.4%
Familiar	305	34.8%
Unknown	178	20.3%

Explore determinants of “Is Donor 2019” -> Yes /No

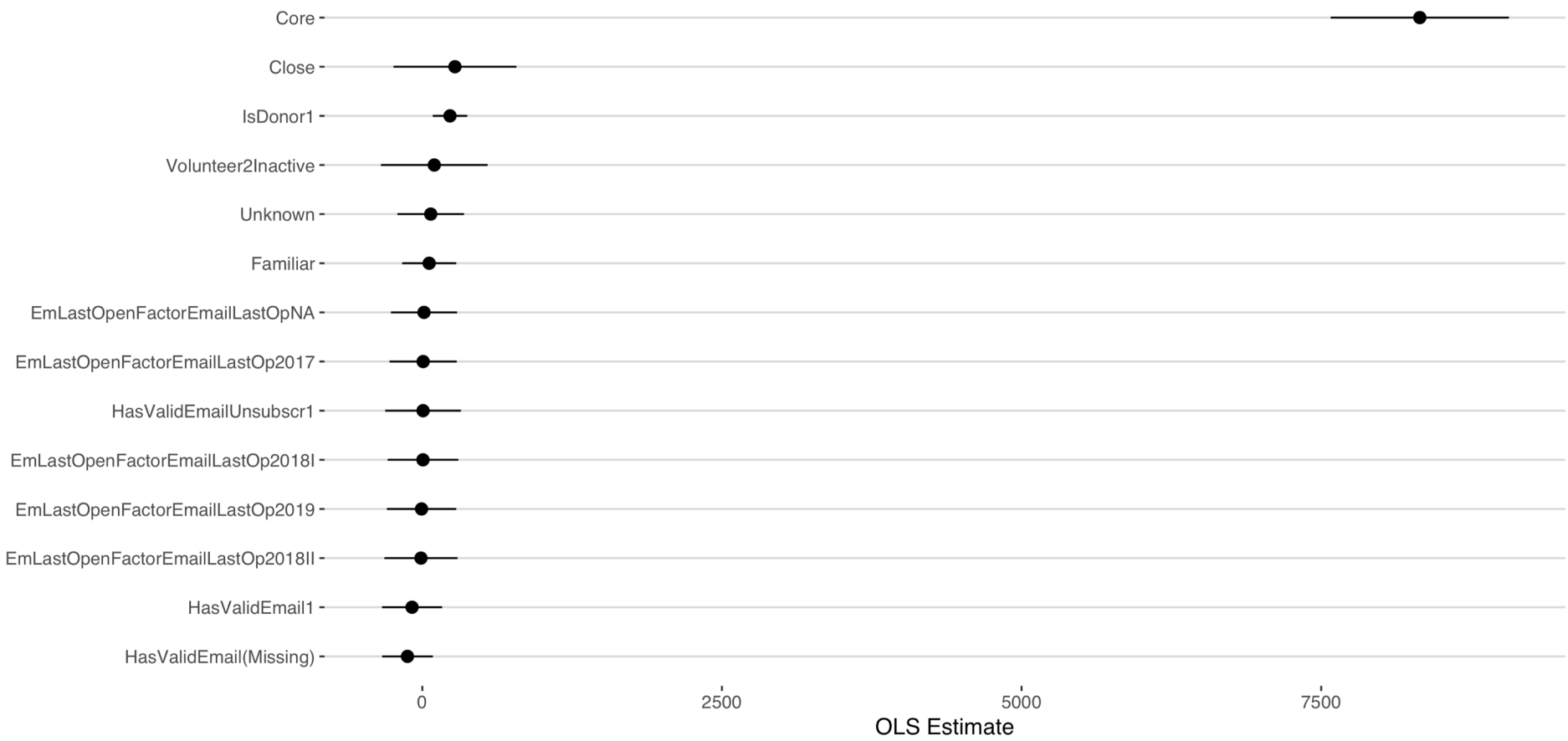
Decision Tree classification model (on test Accuracy = 0.70 & AUC = 0.64)



Linear regression model to estimate "Donation \$ for NYE19"

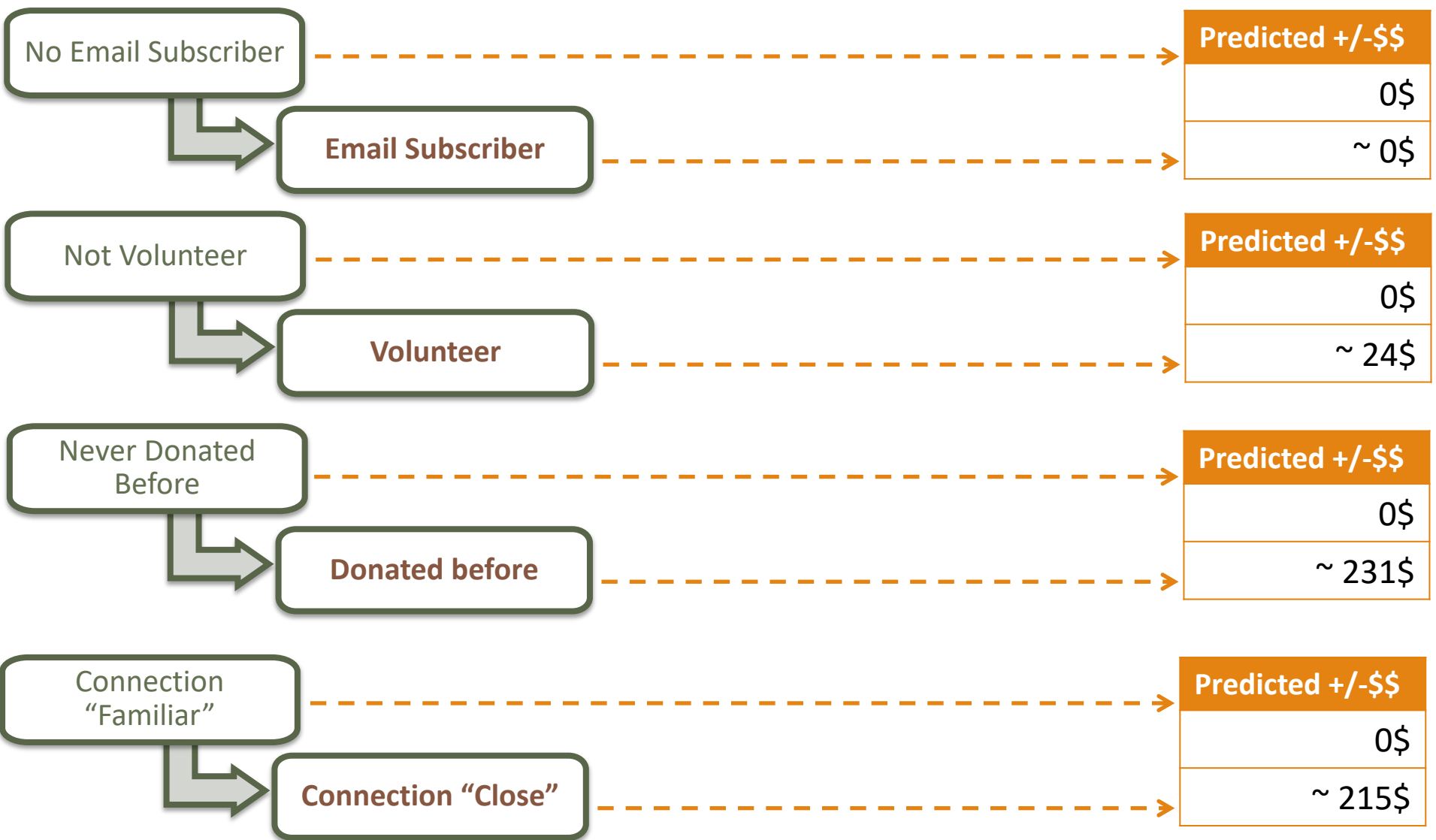
Conf Int of estimates of Lin Mod "lm5_19"

Reference: Connection = Missing; EmLastOpen = bef2017; Volunteer; NotUnsubscr; NotDonor



Created on 2019-03-29

Trying to show the incidence of EACH characteristic as a predictor of \$\$



3) Thoughts and suggestions

Data entry effort urgently needed...

Many interesting information that could help us characterize actual and potential “donors” are actually **missing:**

Who could help?

Variable about Contact	% Missing in Salesforce
Gender	100%
Age	100%
City	75%
Country	95%
Email	11%
Phone	98%
Religion	99%
Marital status	95%
Is Volunteer?	15% (but many are “guessed”)
Connection level	11% (but many are “guessed”)

Bottom line ;)

- 1) Severe lack of important data in Salesforce
 - location? demographic info? (gender, age, household composition)
 - we need volunteers for data entry!!!

- 2) Different sources of fundraising at NYE seem to perform very differently across years (proportional to effort?)
 - Given effort/result, where should we focus this year?

- 3) Core and Closely connected friends are also EXTREMELY generous donors.
 - Worth growing from that...

- 4) Email blasts must be re-discussed and re-focused
 - Albeit challenges of indicators

() IMPORTANT DISCLAIMER:*

The data currently in Salesforce is very incomplete, so this serves as a first exploration and NOT as conclusive analysis...

4) EXTRA

Rebalancing the sample (Theory)

PROBLEM: With highly imbalanced data (only 1.6% donors), *Donors* are so rare that the **?model?** could be biased

ConnectionLevel__c__NA	n	percent
Core	6	2.6%
Close	9	3.8%
Familiar	39	16.6%
Unknown	7	3.0%
(Missing)	174	74.0%

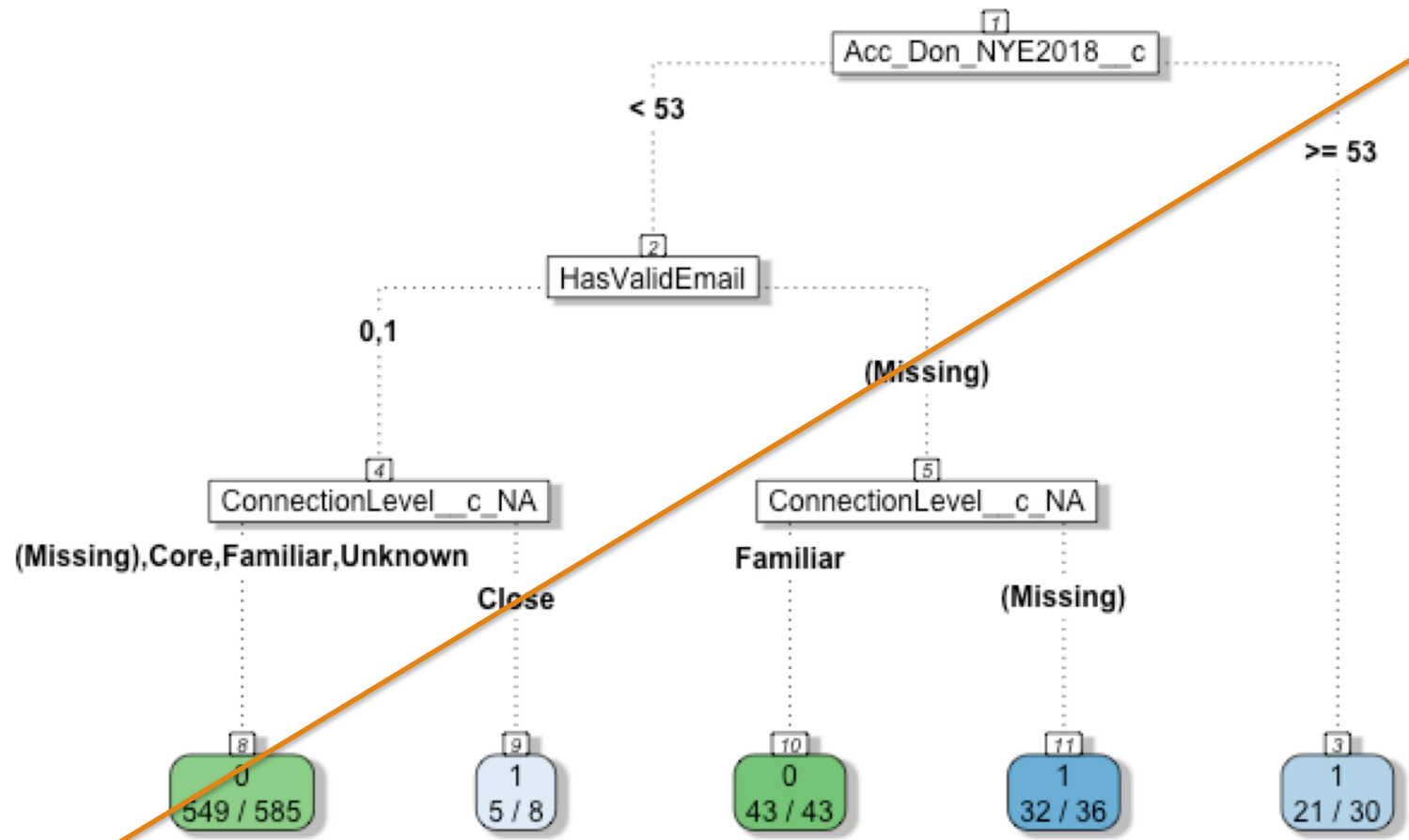
SOLUTION: Of those that are *Not Donors* I take 10% of each level in category “Connection”

- Sample = 887
- Train =
- Test =

ConnectionLevel__c__NA	n	percent
(Missing)	376	42.9%
Core	6	0.7%
Close	12	1.4%
Familiar	305	34.8%
Unknown	178	20.3%

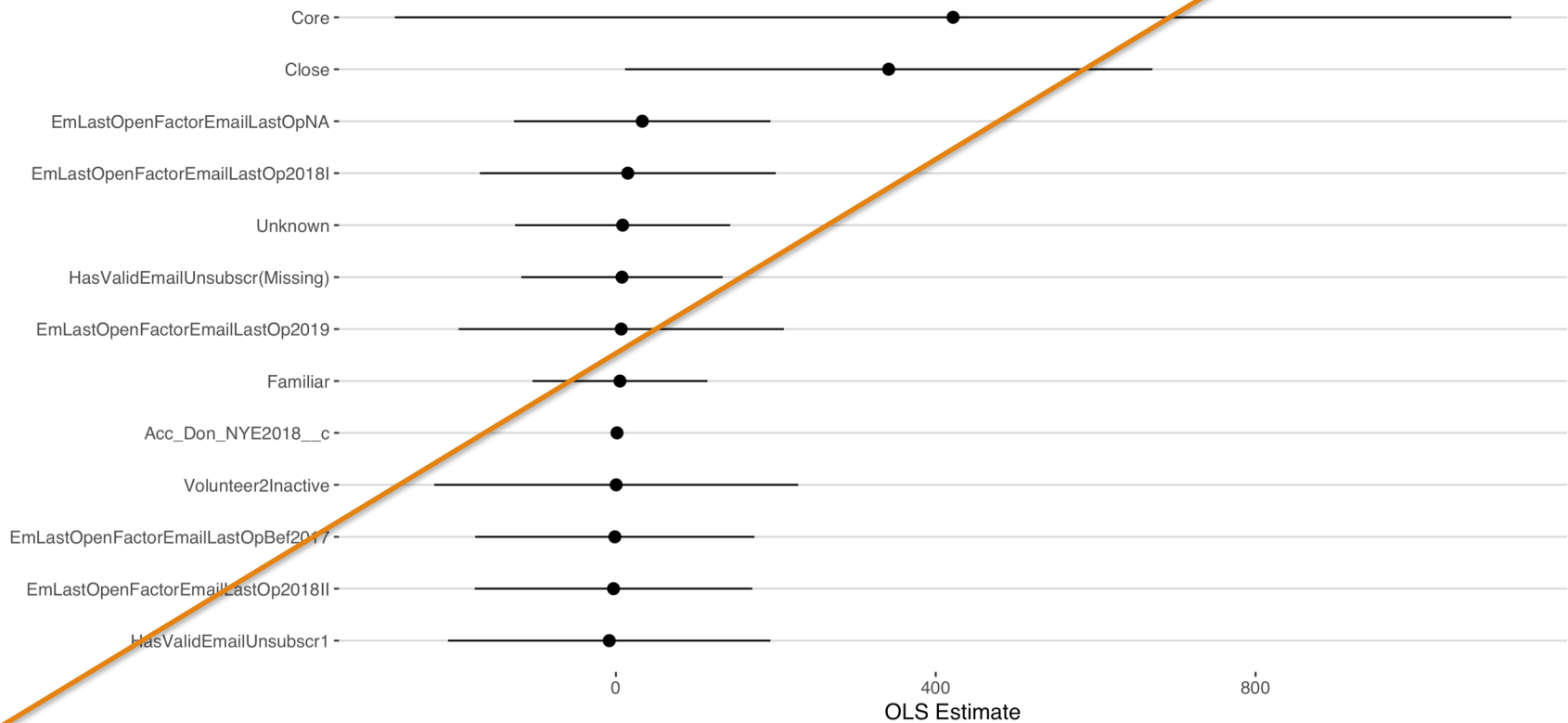
Explore determinants of “Is Donor 2019” -> Yes /No

Decision Tree classification model **WITH** lag donations... (on test Accuracy = 0.85 & AUC = 0.70)



Fitting a linear regression model (with lagged donation 2018) to estimate "Donation \$ for NYE19"

Conf Int of estimates of Lin Mod "lm_ts_6_19"
 Reference: Connection = Missing; EmLastOpen = 2017; Volunteer; NotUnsubscr



Created on 2019-03-27